

Review and Outlook

Machine Learning 2020
mlvu.github.io

the plan

part 1:

Review, Exam strategies

Future work

Causality, Generalization, Compositionality

part 2:

The social impact of machine learning

2

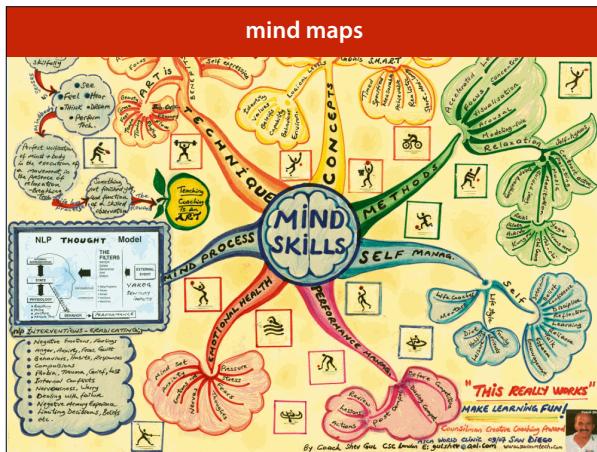


image source: Shev Gul

Mind Skills mind map. Created by Shev Gul

<http://www.mindmapart.com/>

exam strategy

Focus on the LECTURES, not the reading

Read the slides before watching the videos

Focus on the first 10 lectures

Make quick passes over everything. Figure out *what you don't understand*, then move on.

4

studying tricks

Compose a keyword list

Pages > Terminology

Come up with your own exam questions

Make random combinations

5

focus on the ins and outs

$$\begin{aligned}\nabla \mathbb{E}_a r(a) &= \nabla \sum_a p(a)r(a) \\&= \sum_a \nabla p(a)r(a) \\&= \sum_a p(a) \frac{\nabla p(a)}{p(a)} r(a) \quad \nabla \ln(z) = \frac{1}{z} \nabla z \\&= \sum_a p(a) \nabla \ln p(a) r(a) \\&= \mathbb{E}_a r(a) \nabla \ln p(a)\end{aligned}$$

6

Home
Announcements
Syllabus
Pages
Assignments
People
Discussions
Grades
Conferences
Collaborations
Modules
Quizzes
Files
Outcomes
Settings

Recommended reading

These materials are not required to pass the exam. But they are worth looking into if you just want to learn more.

Introduction

- [A few useful things to know about machine learning.](#) ↗ Pedro Domingos.
- [Machine Learning crash course.](#) ↗ From Google.
 - The [glossary](#) ↗ may be particularly useful if you get stuck on an unfamiliar term.

Linear Models 1

- A good way to [visualize squared error loss.](#) ↗

Methodology 1

- Derived features: <https://developers.google.com/machine-learning/crash-course/derived-features>

Methodology 2

Machine Learning Crash Course

OVERVIEW COURSE EXERCISES GLOSSARY

Introduction

Prerequisites and Prework

ML Concepts

- Introduction to ML (3 min)
- Framing (15 min)
- Descending into ML (20 min)
- Reducing Loss (60 min)
- First Steps with TF (60 min)
- Generalization (15 min)
- Training and Test Sets (25 min)
- Validation (40 min)
- Representation (65 min)
- Feature Crosses (70 min)
- Regularization: Simplicity (40 min)
- Logistic Regression (20 min)
- Classification (90 min)
- Regularization: Sparsity (45 min)
- Introduction to Neural Nets (55 min)
- Training Neural Nets (40 min)
- Multi-Class Neural Nets (50 min)
- Embeddings (80 min)

Introduction to Machine Learning

This module introduces Machine Learning (ML).

Estimated Time: 3 minutes

Learning Objectives

- Recognize the practical benefits of mastering machine learning
- Understand the philosophy behind machine learning

Introduction to Machine Learning

Machine Learning Crash Course

OVERVIEW COURSE EXERCISES GLOSSARY

Machine Learning Glossary

This glossary defines general machine learning terms as well as terms specific to TensorFlow.

A

A/B testing

A statistical way of comparing two (or more) techniques, typically an incumbent against a challenger. Testing aims to determine not only which technique performs better but also to understand if the difference is statistically significant. A/B testing usually considers only two techniques for measurement, but it can be applied to any finite number of techniques and measures.

accuracy

The fraction of predictions that a [classification model](#) got right. In [multi-class classification](#), accuracy is defined as follows:

<https://developers.google.com/machine-learning/crash-course/glossary>

Correct Predictions

☰ Seeing Theory

EN

Chapter 1

Basic Probability

This chapter is an introduction to the basic concepts of probability theory.

Chance Events

Expectation

Variance

2018 - P4

All Search by title or author...

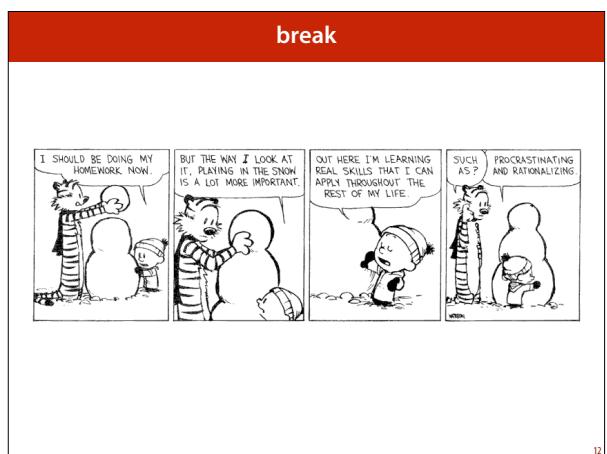
Home Announcements Syllabus Pages Assignments People **Discussions** Grades Files Quizzes Outcomes Collaborations Conferences Modules Settings

▼ Pinned discussions

- Terminology and notation All sections Last post at 24 Feb at 15:59
- Typo and other small mistakes All sections Last post at 7 Mar at 11:31

▼ Discussions

- Questions exam 2018 All sections Last post at 20 Mar at 17:04
- Hw6, Decision Trees question 3 All sections Last post at 18 Mar at 16:26
- Project baseline requirements



what can't we do yet?

Causality

Compositionality

Generalization

13

causality

correlation does not imply causation

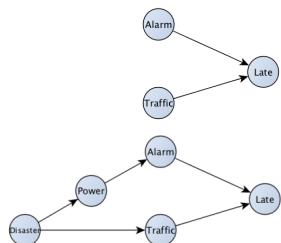
offline learning can only find correlations.

identifying causation requires intervention
i.e. a controlled experiment

14

causality without experiments

background knowledge



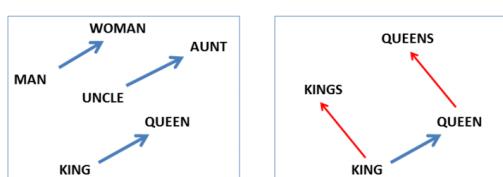
source:<https://medium.com/causal-data-science/if-correlation-doesnt-imply-causation-then-what-does-c74f20d26438>

15

compositionally

$$v(\text{king}) + v(\text{woman}) - v(\text{man}) \approx v(\text{queen})$$

"feminine" vector



16

generalisation

What if your **test data** is a little different from your **training data**.

For instance:

train an RNN to sum numbers between 1 and 10
test on numbers between 1 and 15

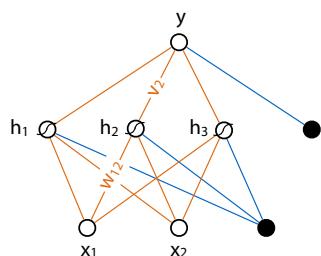
17

causality, compositionality, generalisation

The key is to create a model with the right **inductive bias**

18

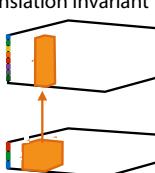
inductive biases: MLP



19

inductive biases of CNNs

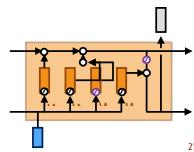
- The data has a grid structure
we know the data consists of pixels
- Inputs far apart on the grid are not relevant for low-level features
we connect only a local group of pixels to each hidden node
- low level feature extractors are translation invariant
we re-use the same weights for each patch



20

inductive biases of LSTMs

- The data is a sequence
- Each token can be modelled as a result of the tokens preceding it.
- Many tokens can be forgotten, and we can infer this from the token itself, together with the immediate context.



inductive bias

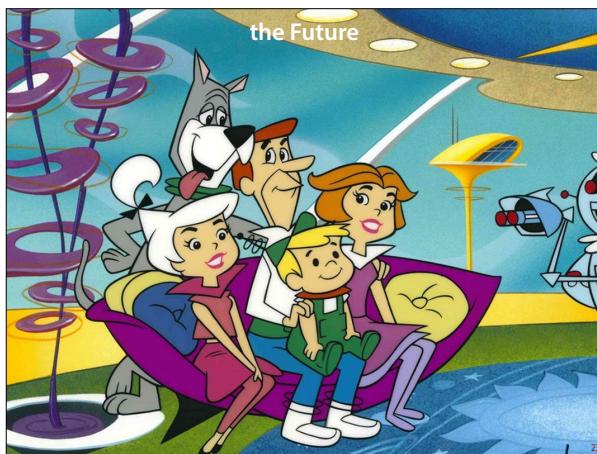
causality: inject background knowledge as an inductive bias

compositionality: add preference for compositionality explicitly, or model the rules of composition

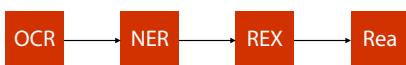
generalization: the more we constrain our model, the better it generalizes
but the less robust it is against the thing we didn't model

Grand challenge: start with the inductive bias, and let the model follow.

22

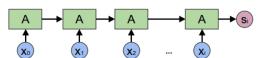


end-to-end learning

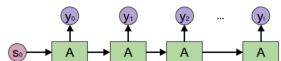


24

software 2.0 / differential programming



fold = Encoding RNN
Haskell: foldl a s



unfold = Generating RNN
Haskell: unfoldr a s

medium.com/@karpathy/software-2-0

colah.github.io/posts/2015-09-NN-Types-FP/

25

Using generative models to invent meaningful creative operations

Let's look at an example where a machine learning model makes a new type of interface possible. To understand the interface, imagine you're a type designer, working on creating a new font¹. After sketching some initial designs, you wish to experiment with bold, italic, and condensed variations. Let's examine a tool to generate and explore such variations, from any initial design. For reasons that will soon be explained the quality of results is quite crude; please bear with us.



distill.pub/2017/aia/

26



Smiling

Images from Sampling Generative Networks by White [4].

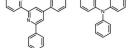
FACES

"It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife."

Length

Sentence from *Pride and Prejudice* by Jane Austen, interpolated by the authors, inspired by experiments done by the novelist Robin Sloan [5].

SENTENCES



Delayed fluorescence decay rate

Images from Automatic chemical design using a data-driven continuous representation of molecules by Gómez-Bombarelli et al [6].

MOLECULES

27

AI
artificial intelligence
IA
intelligence augmentation
II
Intelligent Infrastructure

28

GPT-2

SYSTEM PROMPT (HUMAN-WRITTEN)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

MODEL COMPLETION (MACHINE-WRITTEN, 10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.

Turkish English Spanish Detect language English Spanish Arabic Translate

O bir doktor. O bir hemşire.
x He is a doctor. She is a nurse.

28/5000

 BUSINESS INSIDER Taylor Lorenz Apr 10, 2015, 11:18 AM

The first woman who appears in a Google Image search for 'CEO' is Barbie

The first female result in a Google Image search for "CEO" is Barbie and is hidden at least 10 rows down, a recent post from The Verge points out.

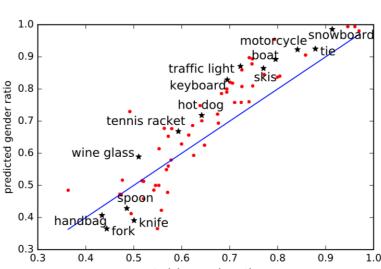
And in fact, it's not even the real Barbie. As T.C. Sottek at the Verge notes, the image of Barbie in a power suit is from a 2005 Onion article stating that "women don't run companies," they just "work behind the scenes to bring a man's vision to light."

Joy Buolamwini

machine learning can amplify data bias



COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	∅
HEAT	STOVE
TOOL	SPATULA
PLACE	KITCHEN



Film fans see red over Netflix 'targeted' posters for black viewers

The streaming service's customers say they are being duped by marketing that shows minor cast members as leading characters



▲ Set It Up is made to look like a two-hander between Taye Diggs and Lucy Liu, rather than the white couple.
Photograph: Twitter Kelly Quantrell @codekelly

For anyone familiar with spending an unrelaxing hour scrolling through the Netflix menu trying to work out what to watch, the idea that one of the [\(20\) netflix-film-black-viewers-personalised-marketing-targeting-imp-1](#) [zonality your viewing choices](#)

Bernard Parker, left, was rated high risk; Dylan Fugitt was rated low risk. [Josh Ritchie for ProPublica]

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

What if

the data is a fair representation of the population

and

the predictions are accurate?

34

racial profiling

TOP STORIES

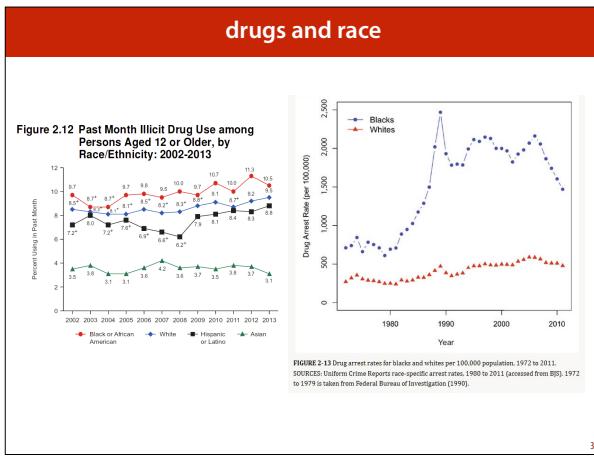
POLICE RACIAL PROFILING OVERWHELMINGLY APPROVED BY DUTCH PUBLIC

By Janene Pieters on June 6, 2016 - 09:02

[https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?
utm_campaign=comms&utm_source=comms-pitch&utm_medium=email&utm_term=algorithm](https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?utm_campaign=comms&utm_source=comms-pitch&utm_medium=email&utm_term=algorithm)

Recently, a Dutch hip-hop artist called Typhoon was stopped by the police. The police admitted that the combination of his skin colour and the fact that he drove an expensive car played a part in the choice to stop him. This caused a small stir in the Dutch media and a nationwide discussion about **racial profiling**, using racial features to predict the likelihood of a person committing a crime. Other examples of profiling include travel security checking people of arabic descent more than others, giving people of certain background higher health-insurance premiums or managers being less likely to hire women for technical positions (gender profiling).

Profiling is an important subject now that machine learning and data mining are becoming more widespread. Since we generally optimise purely for performance, and feed the algorithm lots of features, there is no telling whether it is using sensitive feature like race. An automatic system built to detect whether cars should be stopped for a random search might be very effective at predicting crimes, but unless it's 100% effective, it will stop innocent people to, and it may be engaging in racial profiling.

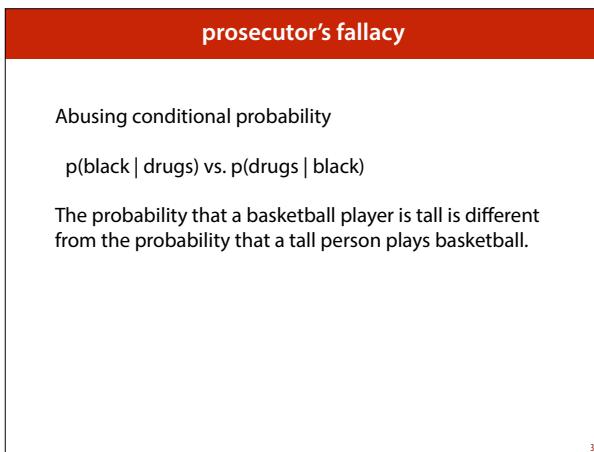


Talking about things like “the probability that a black person commits a crime” is a reductive way to speak, regardless of our intentions, so we will try to make things more concrete with an example: we will look at this recent case, discussed in the Washington Post. In the US, people are routinely classified by ethnicity in research (since it is an important issue) and there are clear guidelines for making the classification. We will follow this definition of a black person. Instead of crime in general, we will focus on the illicit use of drugs. For this, we have good data on how many people engage in illicit drug use and for how many people are arrested for it.

Here we see the rates of illicit drug use broken down by ethnicity, and the arrest rate also broken down by ethnicity. As we see, there is a very small discrepancy between the black/white difference in illicit drug use, ad a huge margin in the difference in arrests.

source:

<http://skeptics.stackexchange.com/questions/36797/do-black-people-and-white-people-use-drugs-at-the-same-rate-in-the-usa-but-blac>
https://www.washingtonpost.com/news/wonk/wp/2013/06/04/the-blackwhite-marijuana-arrest-gap-in-nine-charts/?utm_term=.322fc255f412



Racial profiling is a classic case of the prosecutor’s fallacy. In this case the probability $p(\text{drugs} \mid \text{black})$ is very slightly higher than the probability $p(\text{drugs} \mid \sim\text{black})$, so the police feel that they are justified in using ethnicity as a feature for predicting drug use (it “works”). However, the probability $p(\text{drugs} \mid \text{black})$ many still be very much lower than the probability $p(\sim\text{drugs} \mid \text{black})$ a probability that is never considered. As we see in the previous slide the rates are around $p(\text{drugs} \mid \text{black}) = 0.09$ vs. $p(\sim\text{drugs} \mid \text{black}) = 0.81$. If the police blindly stop only black people, they are disadvantaging over 80% of the people they stop.

This is racism in a nutshell. People are disadvantaged not because of their actions, but because of a feature they share with somebody else who perpetrated a crime. A fundamental property of our modern morality is that people should only be judged by their own actions, and should only be punished for their own actions. Under the banner of “it works” machine learning and data mining can be responsible for horrible acts of discrimination when only their performance for a given task is evaluated and not whether or not their actions are fair.

Assuming that tall people play basketball may work better than assuming that short people play basketball, but you’ll still assume that a lot people play basketball who don’t. If that assumption carries a negative consequence must be taken into account. We end up disadvantaging people purely on the basis of a future

they share with the people whom it is fair to disadvantage.

what if we forbid racial profiling?

Disallow the use of gender, ethnicity, sexual orientation etc. as features in sensitive ML tasks

What about: postcode, hobbies, average salary, mode of transport, etc.

What about companies: how do we police Google, Facebook, Yahoo?

Can we solve the problem by simply disallowing people the use of these sensitive features in datamining applications? The problem is that many other features are highly correlated. If we combine postcode, income, hobbies and music taste (and perhaps the same values of people close by), we can end up with a perfect predictor for the sensitive values. And since the ML algorithm is optimised for what works (in a very narrow sense) rather than what's fair, we will still see algorithms that discriminate.

Moreover, while we can police some institutions, we cannot police foreign companies. If we limit only domestic companies and government institutions, we are just creating an advantage for the institutions we can't control.

What if

the data is a fair representation of the population

and

the predictions are accurate

and

we've correctly used Bayes rule?

actions versus predictions

It is fundamentally unfair to **hold an individual responsible** for the actions of others that share their attributes.

Everybody has the *right* to be judged on their own actions.

"**hold responsible**":

subject to a traffic stop, not give parole, search at an airport, not give a credit card, make it more difficult to get a job.

40

feedback loops

Offline learning doesn't stay offline.

Predictions become actions, that reinforce existing biases from the data.

It's not just about whether the predictions are accurate. It's about whether the **actions are fair, and effective**.

41

Filter Bubble: Breaking Out of the Problematic Loop

the problem of scale



How social media filter bubbles and algorithms influence the election

With Facebook becoming a key electoral battleground, researchers are studying how automated accounts are used to alter political debate online

Revealed: Facebook's internal rules on sex, terrorism and violence

YouTube Updates Recommendations Algorithm to Lessen the Spread of 'Borderline Content'

After complaints that YouTube has essentially given **borderline content** free reign, the video service has announced an update to its algorithm that will rank the highest quality of videos that "serve its users well and can be liked by a majority of viewers in a community." Details below.

Twitter: Algorithms were not always impartial

By Chris Fox Technology reporter

8 September 2018

Children's search terms on YouTube are still skewed towards "disturbing" content. Can anything be done to stem the tide?

Senator Mark Warner of Virginia warns of "optimizing for outrageous, salacious, and often fraudulent content" amid 2016 election concerns.



Opinion Sport Culture Lifestyle

WIRELESS

Technology | Science | Culture | Gear | Business | Politics

China

The complicated truth about China's social credit system

China's social credit system isn't a world first but when it's complete it will be unique. The system isn't just as simple as everyone being given a score though

By NICOLE KOBIE

21 Jan 2019

43

thank you for your attention



mlcourse@peterbloem.nl
